

IMPORTANT FEATURE SELECTION BY SPARSE PCA AND APPLICATION IN AUTOMATED CATEGORIZATION

S. Damla Ahipasaoglu, Ngai-Man Cheung, Peter Richtárik and Martin Takáč
Ahipasaoglu@sutd.edu.sg, N.Man.Cheung@gmail.com, Peter.Richtarik@ed.ac.uk, Takac.MT@gmail.com

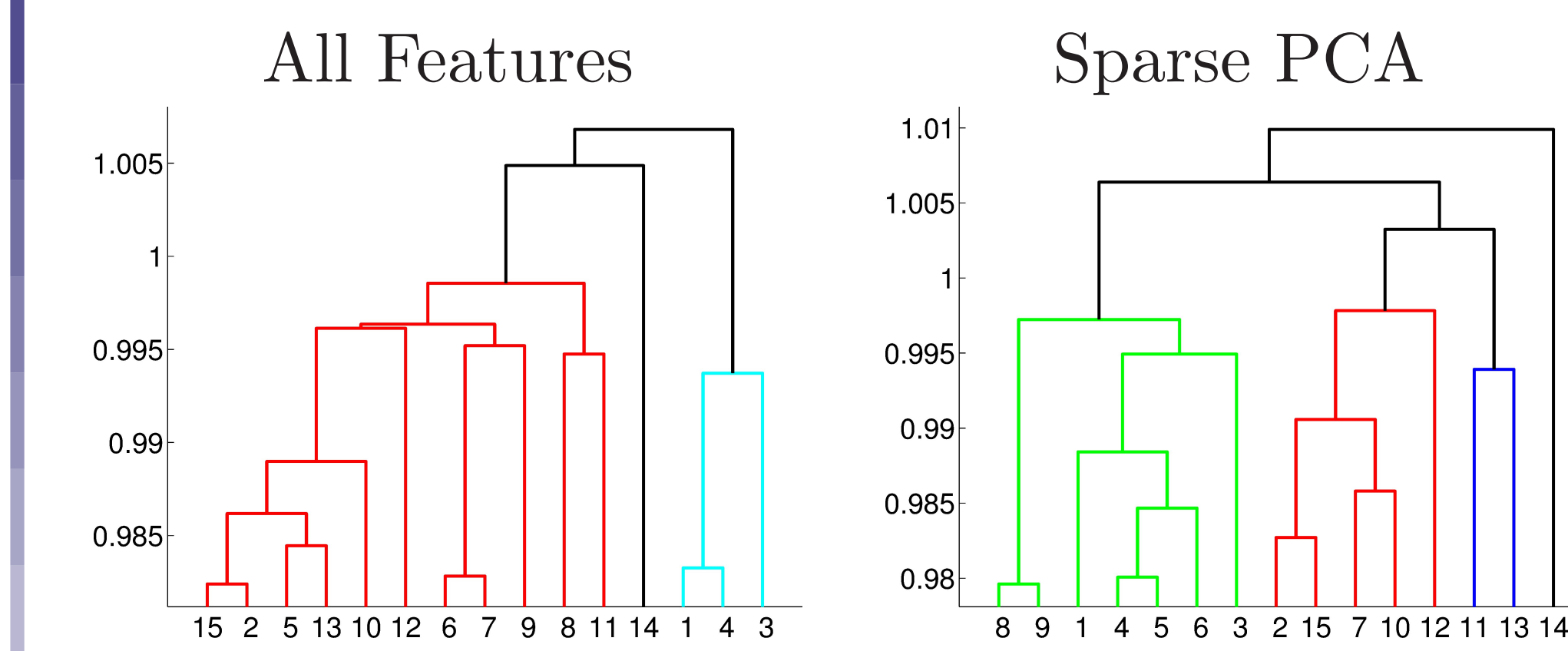
1. INTRODUCTION

Feature - Interesting points (corners, edges, ...) on the picture is equipped with 64 numbers (descriptor)

Drawback - Often unrelated features are extracted
- belonging to noise (people staying in front of building, trees, ...)

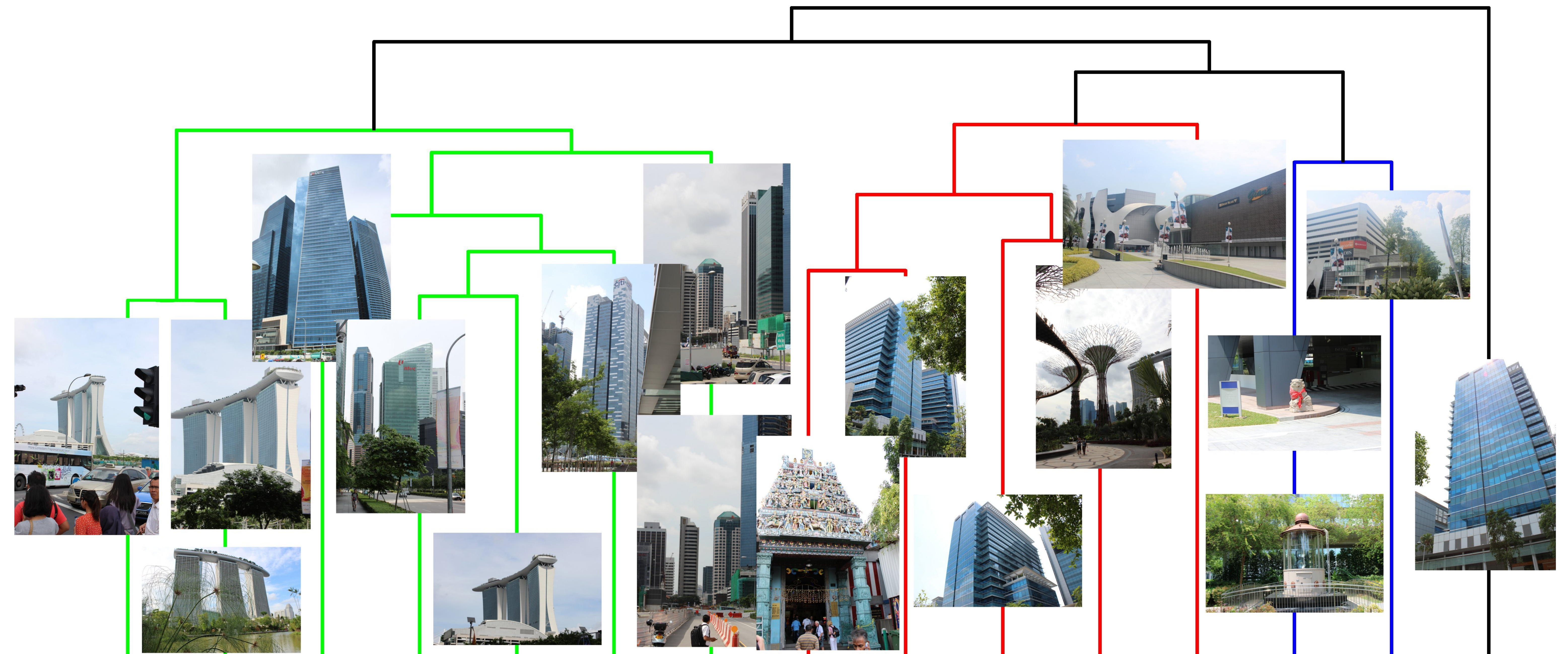
Goal of Good Feature Selection: Identify features, which helps to distinguish objects of different categories

2. HIERARCHICAL CATEGORIZAT.



Features selected by Sparse PCA make it easier and more robust in clustering

4. CATEGORIES IDENTIFIED BY CLUSTERING USING A SUBSET OF FEATURES (SELECTED BY SPARSE PCA)



3. SPARSE PCA IS ABLE TO SELECT IMPORTANT FEATURES



Top - original image, Middle - image with all features (MSRE), Bottom - features selected by sparse PCA

5. FINAL CATEGORIES VS. GROUND TRUTH

Identified Categories	Ground Truth Categories														
	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	19	0	2	0	0	0	0	0	0	0	0	0
0	0	0	0	0	3	0	0	0	0	0	0	3	0	0	0
0	0	0	0	0	0	0	19	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0
0	8	0	0	0	0	0	0	3	8	7	18	6	46	14	0
0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0
0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	0	0	0	0	8	0	0
0	0	0	0	0	0	0	17	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	12
0	0	0	0	0	0	0	0	0	0	17	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0

Total 265 pictures in 15 categories.